CONTENTS

6 Tests of Hypotheses

6.1 Simple Hypothesis

In scientific activities, much attention is devoted to answering questions about the validity of theories or hypotheses concerning physical phenomena.

- Is a new drug effective?
- Does a lot of manufactured items contain an excessive number of defectives?
- Is the mean lifetime of a component at least some specified amount?

The term **hypotheses testing** will refer to the process of trying to decide the truth or falsity of such hypotheses on the basis of experimental evidence.

Definition 1. If $X \sim f(x;\theta)$, a statistical hypothesis is a statement about the distribution of X. If the hypothesis completely specifies $f(x;\theta)$, then it is referred to as a **simple hypothesis**; otherwise it is called **composite**.

Example 1.
A theory proposes that the yield of a certain chemical reaction is normally distributed $X \sim N(\mu, 16)$. Past experience indicates that $\mu = 10$ if a certain mineral is not present, and $\mu = 11$ if the mineral is present. Our experiment would be to take a random sample of size n. On the basis of that sample, state the null hypothesis and the alternative hypothesis.

Definition 2. The **critical region** for a test of hypotheses is the subset of the sample space that corresponds to rejecting the null hypothesis.

In the normal distribution example, $\bar{X}$ is a sufficient statistic for μ, so we may conveniently express the critical region directly in terms of $\bar{X}$, and we will refer to $\bar{X}$ as the test statistic. Because $\mu_1 > \mu_0$ natural form for the critical region in this problem is to let $C = \{(x_1, \ldots, x_n) | \bar{x} \geq c\}$ for some appropriate constant c. That is, we will reject H_0 if $\bar{x} \geq c$, and we will not reject H_0 if $\bar{x} < c$.

There are two possible errors we may make under this procedure. We might reject H_0 when H_0 is true, or we might fail to reject H_0 when H_0 is false. These errors are referred to as follows:

- Type I error: Reject a true H_0.
- Type II error: Fail to reject a false H_0.

We hope to choose a test statistic and a critical region so that we would have a small probability of making these two errors. We will adopt the following notations for these error probabilities:

- $P[\text{Type I error}] = P[TI] = \alpha$
- $P[\text{Type II error}] = P[TII] = \beta$.

Definition 3. For a simple null hypothesis, H_0, the probability of rejecting a true H_0, $\alpha = P[TI]$, is referred to as the **significance level** of the test. For a composite null hypothesis, H_0, the size of the test (or size of the critical region) is the maximum probability of rejecting H_0 when H_0 is true (maximized over the values of the parameter under H_0).

Definition 4. The **power function**, $\pi(\theta)$, of a test of H_0 is the probability of rejecting H_0 when the true value of the parameter is θ.

Example 2.
Suppose $X_i \sim N(\mu, 16)$, $n = 25$, $H_0 : \mu = 10$, $H_1 : \mu = 11$, $\alpha = 5\%$. Find the probability of Type I and Type II errors, α and β;

(a) if the critical region is $\bar{x} \geq 11.316$,

(b) if the critical region is $C_1 = \{(x_1, \ldots, x_n | 10 < \bar{x} < 10.1006\}$.

Example 3.

For Jones' political poll, $n = 15$ voters were sampled. We wish to test $H_0 : p = .5$ against the alternative, $H_1 : p < .5$. The test statistic is X, the number of sampled voters favoring Jones. Calculate α if we select $\{x \leq 2\}$ as the rejection region.

Example 4.

Refer to Example 3. Is our test equally good in protecting us from concluding that Jones is a winner if in fact he will lose? Suppose that he will receive 30% of the votes ($p = .3$). What is the probability β that the sample will erroneously lead us to conclude that H_0 is true and that Jones is going to win?

6.2 Composite Hypotheses

Example 5.

Again, we assume that $X \sim N(\mu, \sigma^2)$, where σ^2 is known, and we wish to test $H_0 : \mu = \mu_0$ against the composite alternative $H_1 : \mu > \mu_0$. For any alternative $\mu_1 > \mu_0$, the critical region should be located on the right hand tail. The critical value c did not depend on the value of μ_1. For a test at significance level α

(a) What is the critical region?

(b) The power function is

(c) What is $\pi(\mu_0)$?

(d) What is $\pi(\mu)$ for $\mu > \mu_0$?

We also may consider a composite null hypothesis. Suppose that we wish to test $H_0 : \mu \leq \mu_0$ against $H_1 : \mu > \mu_0$ and we reject H_0 if $z_0 = \frac{\bar{x}-\mu_0}{\sigma/\sqrt{n}}$ is satisfied. This is still a size α test for the composite null hypothesis. The probability of rejecting H_0 for any $\mu \leq \mu_0$ is $\pi(\mu)$, and $\pi(\mu) \leq \pi(\mu_0) = \alpha$ for $\mu \leq \mu_0$, and thus the size is $\max_{\mu \leq \mu_0} \pi(\mu) = \alpha$. That is, if the critical region is chosen to have size at μ_0. then the Type I error will be less than α for any $\mu < \mu_0$, so the original critical region still is appropriate here. Thus, the a level tests developed for simple null hypotheses often are applicable to the more realistic composite hypotheses, and P[TI] will be no worse than α.

Example 6.

Let $X_1, \ldots, X_{10}$ be a random sample of size n from an exponential distribution, $X_i \sim Exp(\eta, \theta = 20)$. A test of $H_0 : \eta = 13.51$ versus $H_1 : \eta = 22.61$ is desired, based on $X_{1:n}$. A critical region of size $\alpha = 0.01$ is of the form $\{x_{1:n} \geq c\}$.

(a) Find c.

(b) Determine the power of the test under H_1.

6.3 P-Value

There is not always general agreement about how small α should be for rejection of H_0 to constitute strong evidence in support of H_1. Experimenter 1 may consider $\alpha = 0.05$ sufficiently small, while experimenter 2 insists on using $\alpha = 0.01$. Thus, it would be possible for experimenter 1 to reject when experimenter 2 fails to reject, based on the same data. If the experimenters agree to use the same test statistic, then this problem may be overcome by reporting the results of the experiment in terms of the observed size or p-value of the test, which is defined as the smallest size a at which H_0 can be rejected, based on the observed value of the test statistic.

Definition 5. If W is a test statistic, the p-value, or attained significance level, is the smallest level of significance α for which the observed data indicate that the null hypothesis should be rejected.

Example 7.

Refer to example 4, where $n = 15$ voters were sampled. If we wish to test $H_0 : p = .5$ versus $H_1 : p < .5$, using X = the number of voters favoring Jones as our test statistic, what is the p-value if $X = 3$?

6.4 Most Powerful Tests

Let $X_1, \ldots, X_n$ have joint pdf $f(x_1, \ldots, x_n; \theta)$ and consider a critical region C. The notation for the power function corresponding to C is

$$\pi_C(\theta) = P[X_1, \ldots, X_n \in C|\theta]$$

Definition 6. A test of $H_0 : \theta = \theta_0$ versus $H_1 : \theta = \theta_1$ based on a critical region C^* is said to be a most powerful test of size α if

1. $\pi_{C^*}(\theta_0) = \alpha$, and
2. $\pi_{C^*}(\theta_1) \geq \pi_C(\theta_1)$ for any other critical region C of size α [that is, $\pi_c(\theta_0) = \alpha$]

Such a critical region, C^*, is called a most powerful critical region of size α.

Example 8.

Consider the one random variable X that has a binomial distribution with $n = 5$ and $p = \theta$. Let $f(x;\theta)$ denote the pmf of X and let $H_0 : \theta = \frac{1}{2}$ and $H_1 : \theta = \frac{3}{4}$. The following tabulation gives, at points of positive probability density, the values of $f(x;\frac{1}{2})$, $f(x;\frac{3}{4})$, and the ratio $f(x;\frac{1}{2})/f(x;\frac{3}{4})$.

x	0	1	2	3	4	5
$f(x;\frac{1}{2})$	$\frac{1}{32}$	$\frac{5}{32}$	$\frac{10}{32}$	$\frac{10}{32}$	$\frac{5}{32}$	$\frac{1}{32}$
$f(x;\frac{3}{4})$	$\frac{1}{1024}$	$\frac{15}{1024}$	$\frac{90}{1024}$	$\frac{270}{1024}$	$\frac{405}{1024}$	$\frac{243}{1024}$
$\frac{f(x;\frac{1}{2})}{f(x;\frac{3}{4})}$	32	$\frac{32}{3}$	$\frac{32}{9}$	$\frac{32}{27}$	$\frac{32}{81}$	$\frac{32}{243}$

(a) What are the critical regions of size $\alpha = \frac{1}{32}$ for testing H_0 against H_1.

(b) What is the best critical region of testing of $H_0 : \theta = \frac{1}{2}$ vs $H_1 : \theta = \frac{3}{4}$ on $\alpha = \frac{1}{32}$.

Theorem 1. Neyman-Pearson Lemma Suppose that $X_1, \ldots, X_n$ have joint pdf $f(x_1, \ldots, x_n; \theta)$. Let

$$\lambda(x_1, \ldots, x_n; \theta_0, \theta_1) = \frac{f(x_1, \ldots, x_n; \theta_0)}{f(x_1, \ldots, x_n; \theta_1)}$$

and let C* be the set

$$C^* = \{x_1, \ldots, x_n) | \lambda(x_1, \ldots, x_n; \theta_0, \theta_1) \leq k\}$$

where k is a constant such that

$$P[X_1, \ldots, X_n \in C^* | \theta_0] = \alpha$$

Then C^* is a most powerful critical region of size α for testing $H_0 : \theta = \theta_0$ versus $H_1 : \theta = \theta_1$.

Example 9. Consider a random sample of size n from an exponential distribution, $X \sim EXP(\theta)$. Derive the most powerful test of $H_0 : \theta = \theta_0$ against $H_1 : \theta = \theta_1$ where $\theta_1 > \theta_0$ with level of significant α.

Example 10.

Suppose $X \sim Beta(a = 2, b = 2\theta)$. Based on a random sample of size $n = 1$, find the most powerful test of $H_0 : \theta = 3$ against $H_1 : \theta = 2$ with $\alpha = 0.06$, then compute the power of the test for the alternative $\theta = 2$.

Example 11.

Suppose that we have a random sample of four observations from the density function

$$f(x) = \begin{cases} \frac{1}{24\theta^5} x^4 e^{-x/\theta}, & x > 0 \\ 0, & \text{oterwise} \end{cases}$$

Find the rejection region for the most powerful test of

$$H_0 : \theta = \theta_0 \text{ against } H_1 : \theta = \theta_1,$$

assuming that $\theta_1 > \theta_0$.

Example 12.

Below are two probability functions for X, $f(x|1)$ and $f(x|2)$.

	x						
θ	1	2	3	4	5	6	7
2	0.2	0.15	0.1	0.25	0.1	0.05	0.15
1	0.25	0.15	0.25	0.1	0.1	0.05	0.1

Identify a most powerful size = .15 test of $H_0 : \theta = 1$ vs $H_1 : \theta = 2$.

Example 13.

Assume that X is a discrete random variable. Based on an observed value of X, derive the most powerful test of $H_0 : X = GEO(0.05)$ versus $H_1 : X \sim P0I(0.95)$ with $\alpha = 0.0975$. Then find the power of this test under the alternative.

6.5 Uniformly Most Powerful Tests

Definition 7. Let $X_1, \ldots, X_n$ have joint pdf $f(x_1, \ldots, x_n; \theta)$ for $\theta \in \Omega$, and consider hypotheses of the form $H_0 : \theta \in \Omega_0$ versus $H_1 : \theta \in \Omega - \Omega_0$ where Ω_0 is a subset of Ω. A critical region C^*, and the associated test are said to be uniformly most powerful (UMP) of size α if

$$\max_{\theta \in \Omega} \pi_{C^*}(\theta_0) = \alpha$$

and

$$\pi_{C^*}(\theta) \geq \pi_C(\theta)$$

for all $\theta \in \Omega - \Omega_0$ and all critical regions C of size α.

A UMP test often exists in the case of a one-sided composite alternative, and a possible technique for determining a UMP test is first to derive the Neyman- Pearson test for a particular alternative value and then show that the test does not depend on the specific alternative value.

Example 14. Suppose that $X_1, X_2, \ldots, X_n$ constitute a random sample from a normal distribution with known mean μ and unknown variance σ^2. Find the most powerful $\alpha-$level test of $H_0 : \sigma^2 = \sigma_0^2$ versus $H_1 : \sigma^2 = \sigma_1^2$, where $\sigma_1^2 > \sigma_0^2$. Show that this test is equivalent to a χ^2 test. Is the test uniformly most powerful for $H_1 : \sigma^2 > \sigma_0^2$?

Example 15. Refer to example 11, is the test uniformly most powerful for the alternative $\theta > \theta_0$?

Definition 8. A joint pdf $f(\mathbf{x};\theta)$ is said to have a monotone likelihood ratio (MLR) in the statistic $T = t(\mathbf{X})$ if for any two values of the parameter, $\theta_1 < \theta_2$, the ratio $f(\mathbf{x};\theta_2)/f(\mathbf{x};\theta_1)$ depends on $\mathbf{x}$ only through the function $t(\mathbf{x})$, and this ratio is a nondecreasing function of $t(\mathbf{x})$.

Notice that the MLR property also will hold for any increasing function of $t(\mathbf{x})$.

Example 16. Consider a random sample of size n from an exponential distribution, $X \sim EXP(\theta)$. Show that $f(\mathbf{x};\theta)$ has the MLR property in the statistic $T = \sum X$.

The MLR property is useful in deriving UMP tests.

Theorem 2. If a joint pdf $f(\mathbf{x};\theta)$ has a monotone likelihood ratio in the statistic $T = t(\mathbf{x})$ then a UMP test of size for $H_0 : \theta \le \theta_0$ versus $H_1 : \theta > \theta_0$ is to reject H_0 if $t(\mathbf{x}) > k$, where $P[t(\mathbf{X} \ge k|\theta_0] = \alpha$.

The dual problem of testing $H_0 : \theta \ge \theta_0$ versus $H_1 : \theta < \theta_0$ also can be handled by the MLR approach, but the inequalities in the above Theorem should be reversed. Also, if the ratio is a nonincreasing function of $t(\mathbf{x})$, then H_0 of the theorem can be rejected with the inequalities in $t(\mathbf{x})$ reversed.

Example 17.

Let $X_1, \ldots, X_n$ denote a random sample from a Weibull distribution probability density function(p.d.f.)

$$f(x) = \frac{3}{\theta^3} x_i^{3-1} e^{-(x_i/\theta)^3} I_{(0,\infty)}(x), \theta > 0.$$

(a) show that the uniformly most powerful crtical region of size α for testing $H_0 : \theta \geq 10$ versus $H_1 : \theta < 10$ using monotone likelihood ratio(MLR) property is given by $\sum X_i^3 \leq c$, where c is a constant.

(b) Determine the value of c for $\alpha = 0.09$ and $n = 12$. [Note; qchisq(0.91,24) = 33.71]

(c) For the test in (b), find the value of the power for $\theta = 9.7385$. [Note:$pchisq(36.4991, 24) = 0.9509$]

Theorem 3. Suppose that $X_1, \ldots, X_n$, have joint pdf of the form $f(\mathbf{x}; \theta) = c(\theta)h(\mathbf{x})exp[q(\theta)t(\mathbf{x})]$ where $q(\theta)$ is an increasing function of θ.

1. A UMP test of size for $H_0 : \theta \leq \theta_0$ versus $H_1 : \theta > \theta_0$ is to reject H_0 if $t(\mathbf{x}) \geq k$, where $P[t(\mathbf{X}) \geq k|\theta] = \alpha$.
2. A UMP test of size for $H_0 : \theta \geq \theta_0$ versus $H_1 : \theta < \theta_0$ is to reject H_0 if $t(\mathbf{x}) \leq k$, where $P[t(\mathbf{X}) \leq k|\theta] = \alpha$.

Example 18.

Let $X_1, \ldots, X_{30}$ denote a random sample from a Weibull distribution, $X_i \sim WEI(3, \theta)$. Show that a UMP size 0.02 test of $H_0 : \theta \geq 2$ versus $H_1 : \theta < 2$ using Theorem 3 is $\{\sum X_i^3 \leq k\}$, and then determine k.

6.6 Generalized Likelihood Ratio Tests

Definition 9. Let $\mathbf{X} = (X_1, \ldots, X_n)$ where $X_1, \ldots, X_n$ have joint pdf $f(\mathbf{x}; \boldsymbol{\theta})$ for $\theta \in \Omega$ and consider the hypothesis $H_0 : \boldsymbol{\theta} \in \Omega$ versus $H_1 : \boldsymbol{\theta} \in \Omega - \Omega_0$. The **generalized likelihood ratio** (GLR) is defined by

$$\lambda(\mathbf{x}) = \frac{\max\limits_{\theta \in \Omega_0} f(\mathbf{x}; \boldsymbol{\theta})}{\max\limits_{\theta \in \Omega} f(\mathbf{x}; \hat{\boldsymbol{\theta}})} = \frac{f(\mathbf{x}; \hat{\boldsymbol{\theta}}_{\mathbf{0}})}{f(\mathbf{x}; \hat{\boldsymbol{\theta}})}$$

where $\hat{\boldsymbol{\theta}}$ denotes the usual MLE of $\boldsymbol{\theta}$, and $\hat{\boldsymbol{\theta}}_0$ denotes the MLE under the restriction that H_0 is true.

In other words, $\hat{\boldsymbol{\theta}}$ and $\hat{\boldsymbol{\theta}}_{\mathbf{0}}$ are obtained by maximizing $f(\mathbf{x}; \boldsymbol{\theta})$ over Ω and Ω_0, respectively. The generalized likelihood ratio test is to reject H_0 if $\lambda(x) \leq k$, where k is chosen to provide a size α test.

We see that $\lambda(\mathbf{X})$ is a valid test statistic that is not a function of unknown parameters; in many cases the distribution of $\lambda(\mathbf{X})$ is free of parameters, and the exact critical value k can be determined. In some cases the distribution of $\lambda(\mathbf{X})$ under H_0 depends on unknown parameters, and an exact size critical region cannot be determined. If regularity conditions hold, which ensure that the MLEs are asymptotically normally distributed, then it can be shown that the asymptotic distribution of $\lambda(\mathbf{X})$ is free of parameters, and an approximate size α test will be available for large n. In particular, if $\mathbf{X} \sim f(\mathbf{x}; \theta_1, \ldots, \theta_k),$, then under $H_0 : (\theta_1, \ldots, \theta_r) = (\theta_{10}, \ldots, \theta_{r0}), r < k$, approximately, for large nC,

$$-2 \ln \lambda(\mathbf{X}) \sim \chi^2(r)$$

Thus an approximate size test is to reject H_0 if

$$-2 \ln \lambda(\mathbf{X}) \geq \chi^2_{1-\alpha}(r)$$

Example 19.

Suppose that $X \sim N(\mu, \sigma^2)$ where σ^2 is known and we wish to test $H_0 : \mu = \mu_0$ versus $H_1 : \mu \neq \mu_0$, derive the GLR test.

Example 20.
Suppose that $X \sim N(\mu, \sigma^2)$ where μ ad σ^2 are both unknown and we wish to test $H_0 : \mu = \mu_0$ against $H_1 : \mu \neq \mu_0$, derive the GLR test.

Example 21.
Let $X_1, \ldots, X_n$ be a random sample from an exponential distribution, $X_i \sim EXP(\theta)$. Derive the generalized likelihood ratio (GLR) test of $\theta = \theta_0$ versus $\theta \neq \theta_0$. Determine and approximate critical value for size α using the large-sample chi-square approximation.

Example 22.

Let $X_1, \ldots, X_n$ be a random sample from an exponential distribution, $X_i \sim EXP(\theta)$. Derive the generalized likelihood ratio (GLR) test of $H_0 : \theta = \theta_0$ versus $H_1 : \theta > \theta_0$.

Example 23.

Let $X_1, X_2, \ldots, X_k$ denote a random sample from a gamma distribution $X_i \sim GAM(\alpha_1 = 2, \beta_1)$ and let $Y_1, Y_2, \ldots, Y_l$ denote an independent random sample from a gamma distribution $Y_i \sim GAM(\alpha_2 = 2, \beta_2)$.

(a) Find the likelihood ratio criterion for testing $H_0 : \beta_1 = \beta_2$ versus $H_1 : \beta_1 \neq \beta_2$

(b) Show that the test in part (a) can be based on the statistic

$$T = \frac{\sum X_i}{\sum X_i + \sum Y_i}.$$

6.7 Bayesian Tests of Hypotheses

Tests of hypotheses can also be approached from a Bayesian perspective. When testing $H_0 : \theta \in \Omega_0$ versus $H_1 : \theta \in \Omega_1$. We compute the posterior probabilitis $P^*(\theta \in \Omega_0)$ and $P^*(\theta \in \Omega_1)$ and reject the hypothesis with lower posterior probability. That is for testing $H_0 : \theta \in \Omega_0$ versus $H_1 : \theta \in \Omega_1$, we reject H_0 if $P^*(\theta \in \Omega_0) < P^*(\theta \in \Omega_1)$.

Example 24.

Suppose that X has pdf $f(x|\theta) = 2\theta(1-2x)+2x$ on $[0,1]$. A Bayesian wants to test $H_0 : \theta \leq 0.63$ versus $H_1 : \theta > 0.63$. If the Bayesian prior distribution is $U[0,1]$, what is the person's test?

Example 25.

If $X \sim POI(\lambda)$ and a Bayesian uses a prior for λ that is Exponential with mean 1, what is the Bayes test of $H_0 : \lambda \leq 2$ versus $H_1 : \lambda > 2$?